

➤ **Plant Phenomics data management challenges and solutions based on the Minimum Information About Plant Phenotyping Experiments (MIAPPE)**

Solutions from INRAE as well as ELIXIR and EMPHASIS European Infrastructures and beyond



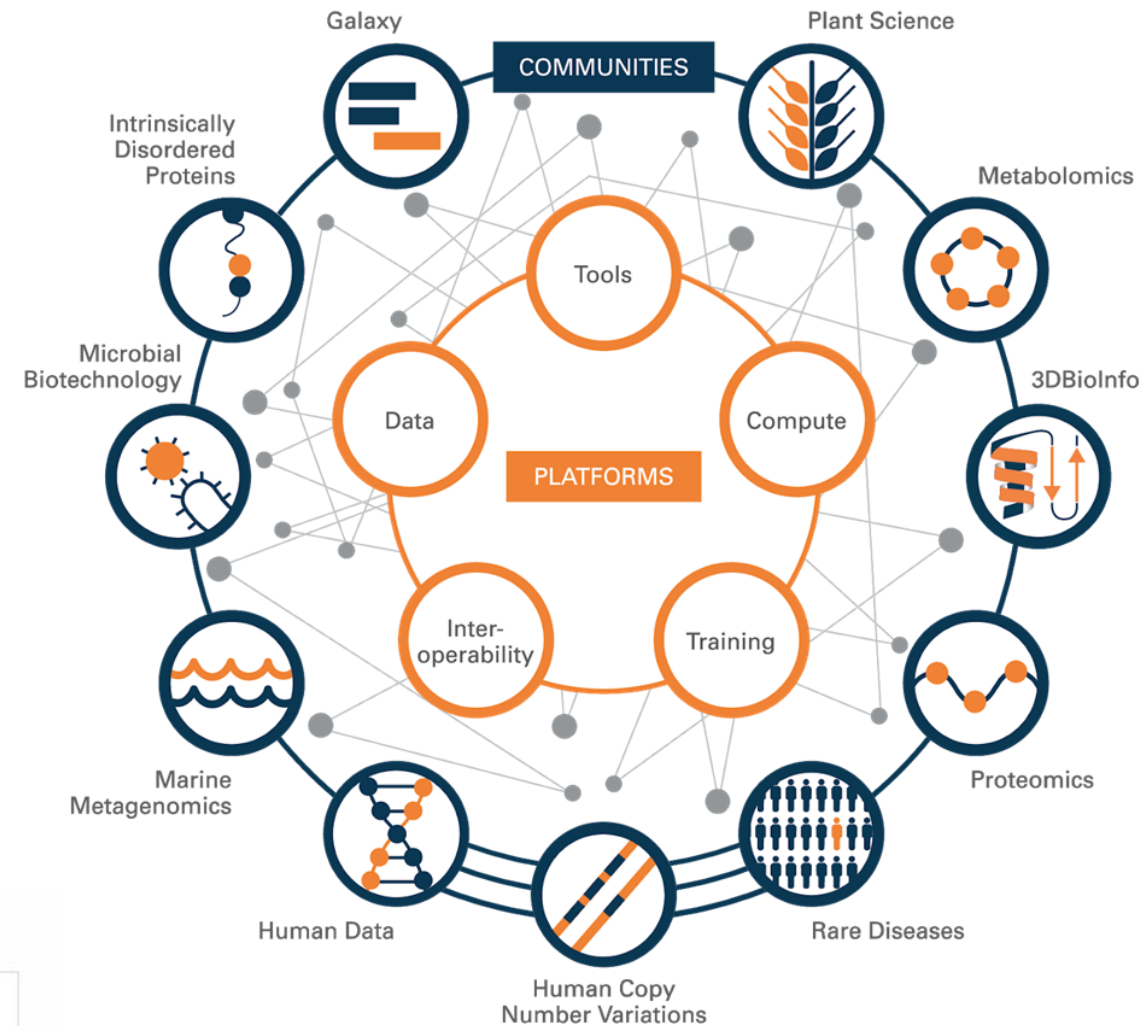
EMPHASIS

[www.elixir-europe.org](http://www.elixir-europe.org)

<https://emphasis.plant-phenotyping.eu>

# ELIXIR Plant community

- European RI for Life science data
- Sustainable tool federation
- FAIR data management
  - Adoption of standards in plant sciences
  - Develop community recommendations
  - Data portal
- Facilitate data integration and analysis
- Training.
- Joint projects (ELIXIR, EU, ...)



## Plant Sciences Community

### Leadership



Sebastian Beier  
(ELIXIR Germany)



Kristina Gruden  
(ELIXIR Slovenia)



Cyril Pommier  
(ELIXIR France)



Katharina Heil  
(Communities Coordinator,  
ELIXIR Hub)





# EMPHASIS

## Infrastructure Categories

PLANT PHENOTYPING REQUIRES INTEGRATION OF BOTH FACILITIES AND ACTIVITIES



CONTROLLED CONDITIONS

Investigation of diverse plant traits in response to well-defined environmental conditions



INTENSIVE FIELD

Detailed investigation of plants and canopies under well-monitored field conditions



LEAN FIELD

Field sites with basic equipment and environmental monitoring that can be linked to a network of field sites



MODELLING

Models integrated in phenotyping pipelines and predictive models using phenotypic data



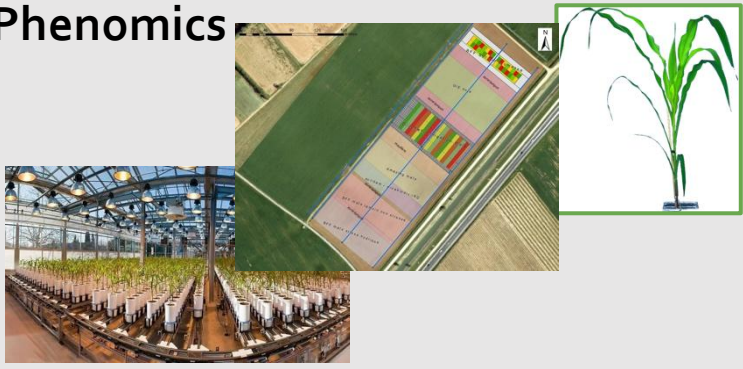
DATA & COMPUTATIONAL SERVICES

Integrating compatible information systems to provide access to data

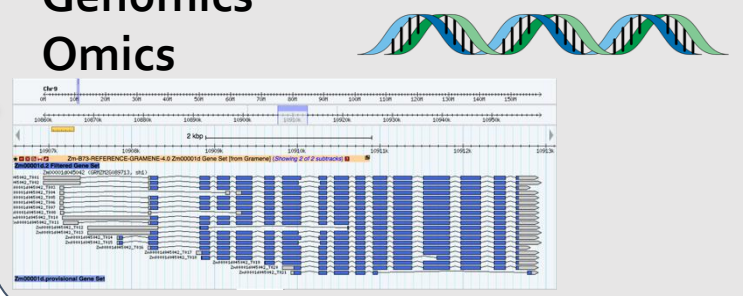


# Plant Phenotype Publication Challenges

## Phenomics



## Genetics Genomics Omics



F<sub>indable</sub> A<sub>ccessible</sub> I<sub>nteroperable</sub> R<sub>eusable</sub>

**Plant Breeding**  
Genetic variations by Traits

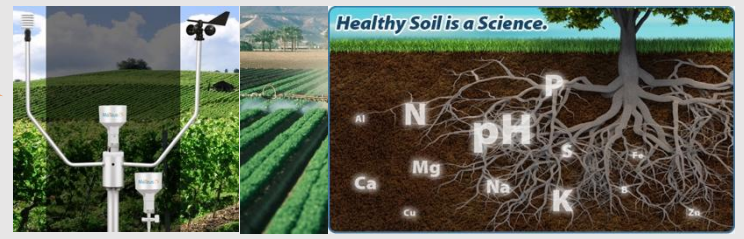
**Climate Change Studies**  
Genotype by Environment

Dispersed  
Heterogenous  
Getting Standardized

Mostly centralized  
Homogenous data  
Heterogenous metadata

F<sub>indable</sub> A<sub>ccessible</sub> I<sub>nteroperable</sub> R<sub>eusable</sub>

## Environment

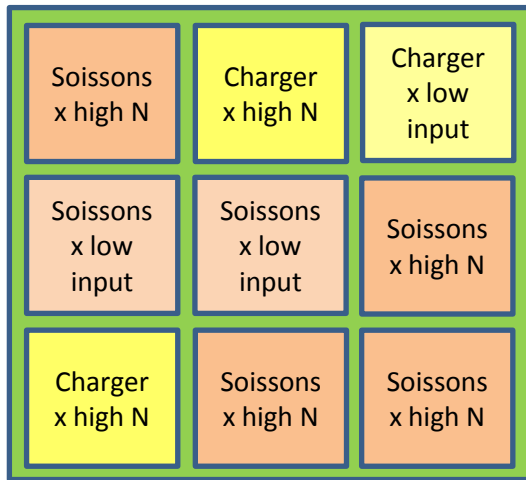


Dispersed  
Heterogenous

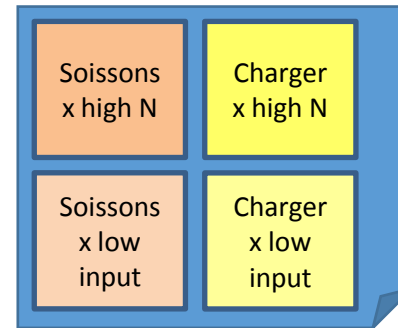
# ● Phenotype data lifecycle



« Raw » data, pheno/env measurement, variables



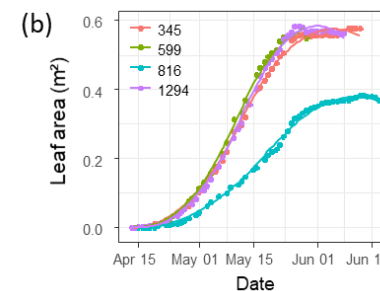
« Derived » data, computed, reduced, indicators



Derivation, Reduction

Genotype	Treatment	N input	Date	Rep	Fusariose
Soissons	low input	15,3225129	15/11/2011	1	5
Soissons	low input	15,3430556	16/11/2011	2	7

Genotyp	Treatment	Fusariose
e		
Soissons	low input	6



Plant Phenomics data management challenges and solutions based on MIAPPE

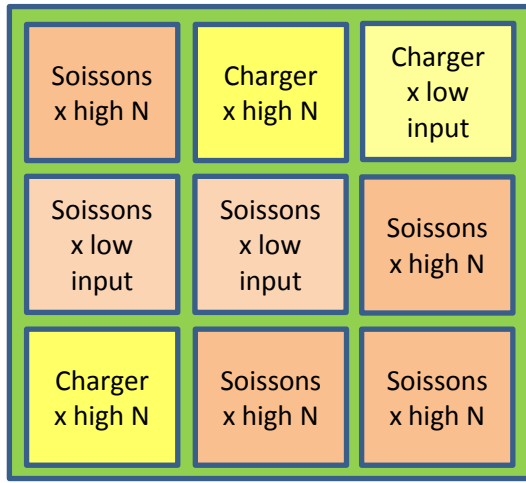
24 October 2023 / Cyril Pommier



# Phenotype data lifecycle

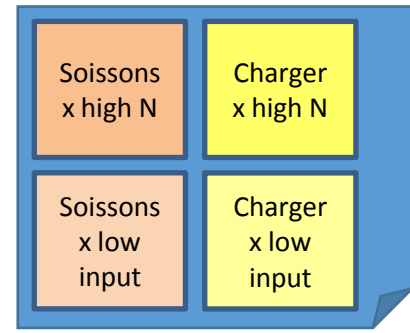


« Raw » data, pheno/env measurement, variables

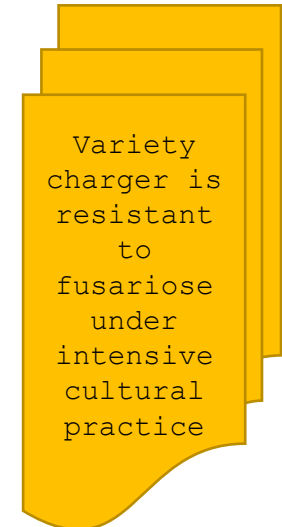


Derivation, Reduction

« Derived » data, computed, reduced, indicators

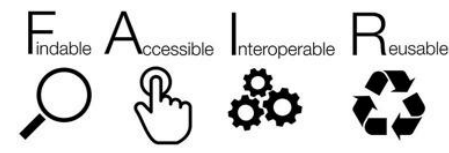
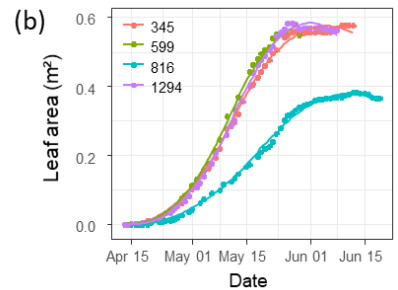


Publication



Genotype	Treatment	N input	Date	Rep	Fusariose
Soissons	low input	15,3225129	15/11/2011	1	5
Soissons	low input	15,3430556	16/11/2011	2	7

Genotyp	Treatment	Fusariose
Soissons	low input	6

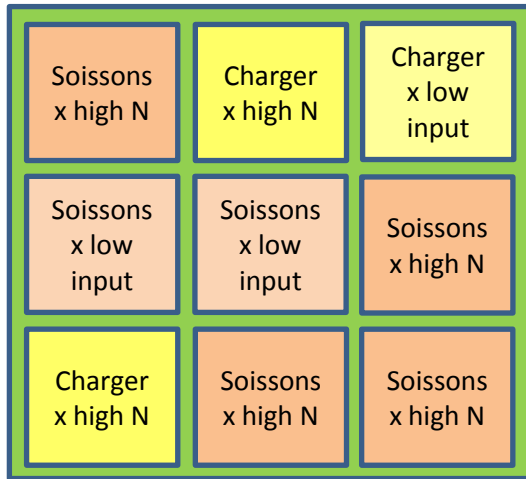


Wilkinson et al.  
**The FAIR Guiding Principles for scientific data management and stewardship.**  
*Scientific Data 3 (2016)*

# Phenotype data lifecycle

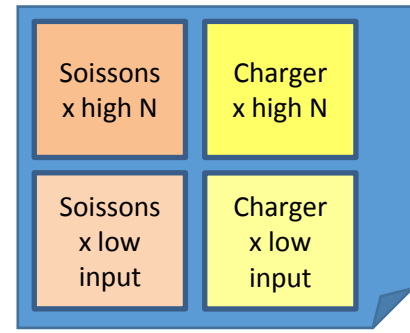


« Raw » data, pheno/env measurement, variables

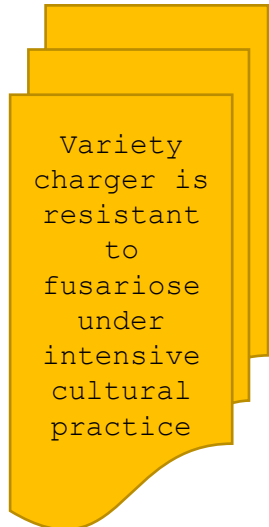


Derivation, Reduction

« Derived » data, computed, reduced, indicators



Publication



Genotype	Treatment	N input	Date	Rep	Fusariose
Soissons	low input	15,3225129	15/11/2011	1	5
Soissons	low input	15,3430556	16/11/2011	2	7

## VARIABLES

- Raw Measures Pheno & Env
- Data cleaning
- Traceability, Reproducibility & Provenance

Genotyp e	Treatment	Fusariose
Soissons	low input	6

## INDICATORS

- New computation for each scientific question
- One experimental dataset → many derived dataset

## DESCRIPTEURS PRIMAIRES

- Caractérisation
- Phenodes
- Accessions

# Phenotype data lifecycle: tools and solutions





# • Sharing standards: standards registries



FAIRsharing.org standards, databases, policies search through all content Q SEARCH LOGIN ↗







STANDARDS DATABASES POLICIES COLLECTIONS ORGANISATIONS ADD CONTENT STATS

ACTIONS ▾

**GENERAL INFORMATION**

This record replaces or incorporates the following deprecated resources:  
[Wheat Data Interoperability Guidelines](#)

  **Plant research data resources for european infrastructures**

Type	Collection
Registry	Collection
Description	Collection of standards, databases, and data resources used for data management in plant research. This collection has been built by the ELIXIR Europe plant community in collaboration with companion infrastructures such as EMPHASIS in particular. It encompasses resources for phenomics to genomic, genetic and other omics. This is tightly connected to the Plant page of the RDM Kit.
Organisations	<a href="#">ELIXIR Europe</a> , <a href="#">Rothamsted Research, UK</a> , <a href="#">VIB-UGent Center for Plant Systems Biology, Ghent, Belgium</a> , <a href="#">BioData.pt</a> , <a href="#">Institut de Recherche pour le Developpement</a> , <a href="#">EMPHASIS</a> , <a href="#">ELIXIR Italy</a> , <a href="#">National Research Institute for Agriculture, Food and Environment</a> , <a href="#">International Cooperation Center for Agricultural Research for Development (CIRAD), France</a> , <a href="#">European Bioinformatics Institute</a> , <a href="#">Wellcome Genome Campus</a>
Homepage	<a href="https://elixir-europe.org/communities/plant-sciences">https://elixir-europe.org/communities/plant-sciences</a>

# Community driven recommendations and registries

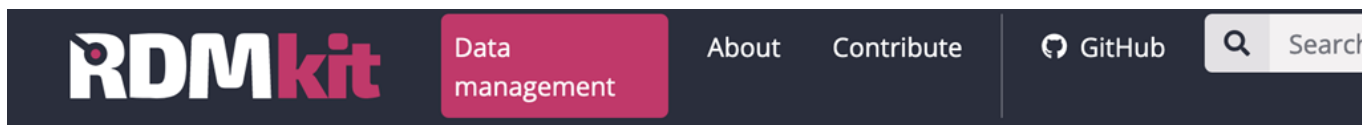
- WheatIS:

<http://wheatis.org/DataStandards.php>












- RDM Toolkit

  - <https://rdmkit.elixir-europe.org/>

- Community story



ARTICLE  
Developing data interoperability using standards:  
a community use case [version 2; referees: 2]

Yeumo<sup>1</sup>, Michael Alaux <sup>2</sup>, Elizabeth Arnaud<sup>3</sup>, Sophie Aubin<sup>1</sup>, Ute Baumann<sup>4</sup>,  
the<sup>5</sup>, Laurel Cooper <sup>6</sup>, Hanna Ówiek-Kupczyńska<sup>7</sup>, Robert P. Davey <sup>8</sup>,  
dan Fulss<sup>9</sup>, Clement Jonquet <sup>10,11</sup>, Marie-Angélique Laporte<sup>3</sup>, Pierre Larmande <sup>12,13</sup>,  
nier <sup>2</sup>, Vassilis Protonotarios <sup>14</sup>, Carmen Reverte <sup>15</sup>, Rosemary Shrestha<sup>9</sup>,  
rats<sup>16</sup>, Aravind Venkatesan <sup>12</sup>, Alex Whan<sup>17</sup>,  Hadi Quesneville <sup>2</sup>

details  
This article is included in the [Global Open Data for Agriculture and Nutrition gateway](#).

## Are you working with data in the Life Sciences? Do you feel overwhelmed when you think about Research Data Management?

The ELIXIR Research Data Management Kit (RDMkit) is an online guide containing management practices applicable to research projects from the beginning to the end of the data life cycle. RDMkit supports FAIR data — Findable, Accessible, Interoperable, Reusable — by-design, from the first steps of data management planning to the final storage of data in public archives.

The RDMkit organises information into the six sections displayed below, which are interconnected but can be browsed independently.

### Data management

Data life cycle

Your role

Your domain

Your tasks

Tool assembly

National resources

All tools and resources

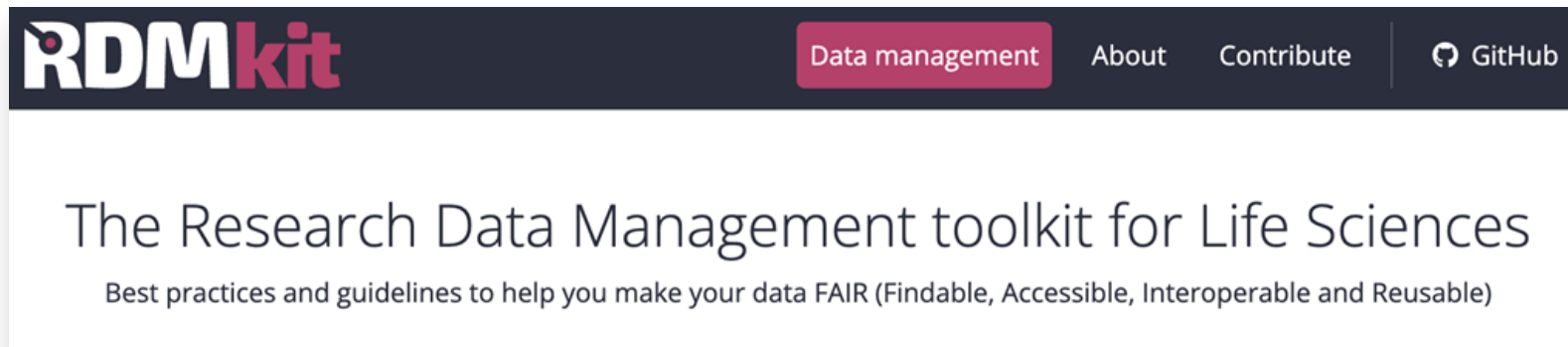
All training resources

## Data life cycle



# Community guidelines portal : RDMkit - Best practices and guidelines for FAIR data management

- A “wikipedia-like” knowledge base website, free and open
- Describes how to manage research outputs according to FAIR principles
- Portal to other online resources used by RDM professionals and researchers



URL: <https://rdmkit.elixir-europe.org/>

Recommended in the **Horizon Europe Program Guide** as the “resource for Data Management guidelines and good practices for the Life Sciences”

# RDMkit is NOT

- A peer reviewed journal
- A repository to store files
- A registry for listing resources and tools
- A manual or user guide for tools
- A website for projects to share deliverables



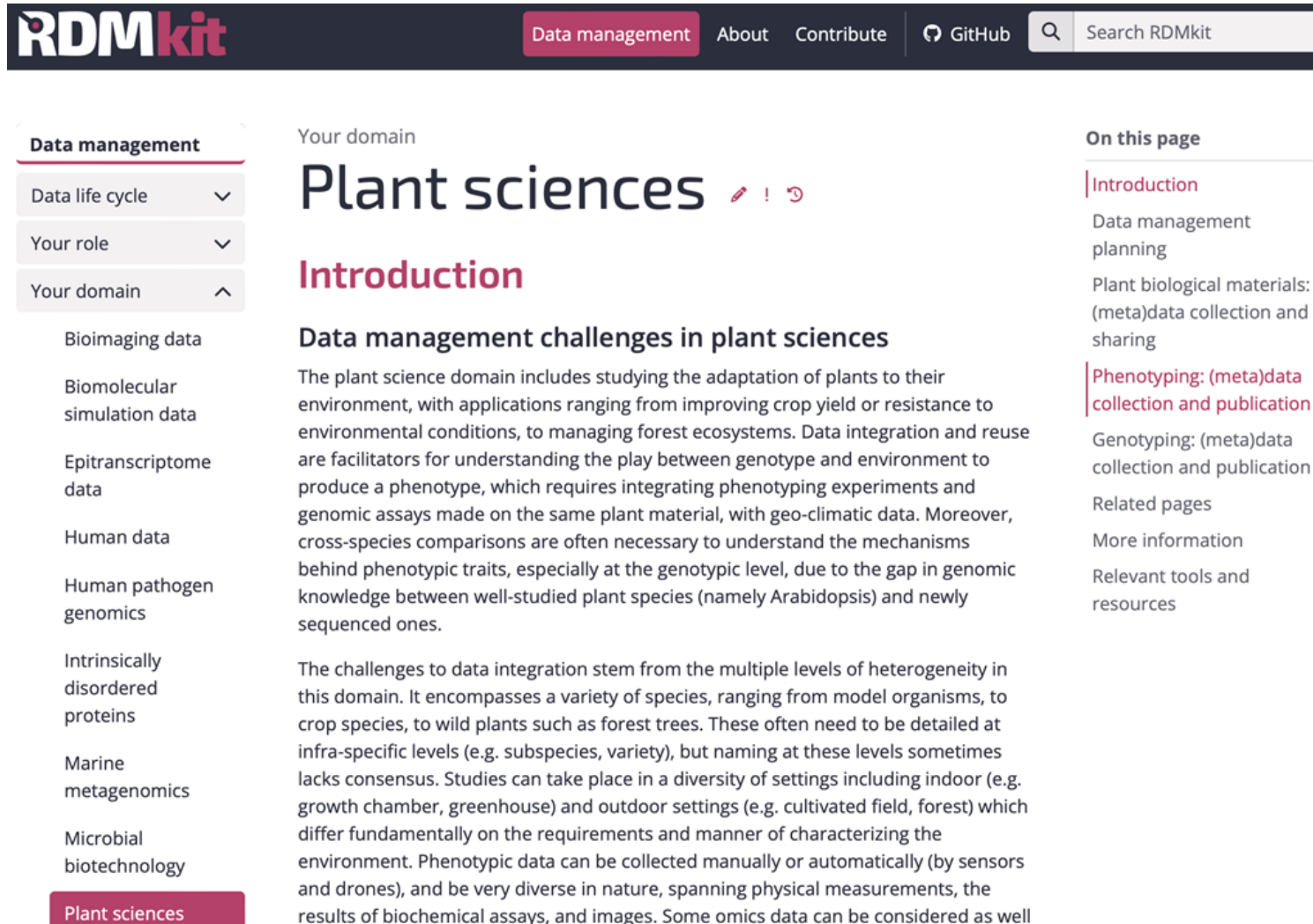
# RDMkit in numbers



- Contributors are experts in RDM and/or in scientific domain in ELIXIR and beyond
- ELIXIR Node service in 3 Nodes/Countries
- Managed by an editorial board composed of 16 ELIXIR members from several Nodes and the lead for “FAIR Data & Resources” at the Office of Data Science Strategy at NIH
- The content is curated by members of the ELIXIR RDM Community

# Plant Pages

[https://rdmkit.elixir-europe.org/plant\\_sciences](https://rdmkit.elixir-europe.org/plant_sciences)



**RDMkit** Data management About Contribute GitHub Search RDMkit

**Data management**

- Data life cycle
- Your role
- Your domain

Bioimaging data

Biomolecular simulation data

Epitranscriptome data

Human data

Human pathogen genomics

Intrinsically disordered proteins

Marine metagenomics

Microbial biotechnology

**Plant sciences**

Your domain

## Plant sciences

### Introduction

#### Data management challenges in plant sciences

The plant science domain includes studying the adaptation of plants to their environment, with applications ranging from improving crop yield or resistance to environmental conditions, to managing forest ecosystems. Data integration and reuse are facilitators for understanding the play between genotype and environment to produce a phenotype, which requires integrating phenotyping experiments and genomic assays made on the same plant material, with geo-climatic data. Moreover, cross-species comparisons are often necessary to understand the mechanisms behind phenotypic traits, especially at the genotypic level, due to the gap in genomic knowledge between well-studied plant species (namely Arabidopsis) and newly sequenced ones.

The challenges to data integration stem from the multiple levels of heterogeneity in this domain. It encompasses a variety of species, ranging from model organisms, to crop species, to wild plants such as forest trees. These often need to be detailed at infra-specific levels (e.g. subspecies, variety), but naming at these levels sometimes lacks consensus. Studies can take place in a diversity of settings including indoor (e.g. growth chamber, greenhouse) and outdoor settings (e.g. cultivated field, forest) which differ fundamentally on the requirements and manner of characterizing the environment. Phenotypic data can be collected manually or automatically (by sensors and drones), and be very diverse in nature, spanning physical measurements, the results of biochemical assays, and images. Some omics data can be considered as well

**On this page**

- Introduction
- Data management planning
- Plant biological materials: (meta)data collection and sharing
- Phenotyping: (meta)data collection and publication
- Genotyping: (meta)data collection and publication
- Related pages
- More information
- Relevant tools and resources



- **PLANT DATA STANDARDS**



# Plant Data standards

## Semantic

- ◆ Description of the data
- ◆ Controlled vocabularies: term name and definitions
- ◆ Ontologies: semantic links between terms
- ◆ *Biologist driven*



## Persistent Unique Identifiers

URI, gene ID, accessions ID, Trait ID, DOI,...

## Structure

- Formatting and Organizing the data
- Data Models
- Standards : CSV, VCF, GFF, MIAPPE ([www.miappe.org](http://www.miappe.org)), etc...
- *Biologist & Computer scientist driven*



## Technical

- Data integration and sharing
- Interoperability : tools and systems
  - GA4GH
  - Breeding API [www.brapi.org](http://www.brapi.org)
  - Computer scientist driven



# • Semantic Standard: Ontologies

## • Annotating one object

- Protein, gene
- Plant, plant anatomy, ...

UniProtKB - P56761 (PSBD\_ARATH)

Display [Help video](#) [BLAST](#) [Align](#) [Format](#) [Add to basket](#) [History](#)

Entry

Publications

Feature viewer

Feature table

None

Function!

Protein **Photosystem II D2 protein**

Gene **psbD**

Organism *Arabidopsis thaliana (Mouse-ear cress)*

Status Reviewed - Annotation score: ●●●●●● - E:

## • Atomic concept

- protein function
- cellular localization
- ...

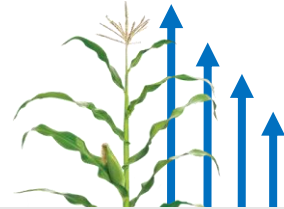
Gene Product	Symbol	Qualifier	GO Term	Evidence	Reference	With / From	Taxon
UniProtKB:P56761	psbD	is_active_in	GO:0009535	ECO:0000318 IBA	PMID:21873635	PANTHER:PTN002108145 more...	3702 Arabidopsis thaliana
UniProtKB:P56761	psbD	part_of	GO:0009523	ECO:0000318 IBA	PMID:21873635	PANTHER:PTN002108145 more...	3702 Arabidopsis thaliana
UniProtKB:P56761	psbD	enables	GO:0005515	ECO:0000353 IPI	PMID:25846821	UniProtKB:Q9FL44	3702 Arabidopsis thaliana
UniProtKB:P56761	psbD	involved_in	GO:0019684	ECO:0000256 IEA	GO_REF:0000002	InterPro:IPR000484 more...	3702 Arabidopsis thaliana

- Semantic Standard: Ontologies for Phenotype

- Describing traits/features in specific plant species
- Crop Ontology Trait + Method + Scale Semantic model



Variable identification: Plant height example



- M1: Total height
- M2: First tassel branch
- M3: Last expanded leaf
- M4: Youngest growing leaf

...There is an uncountable number of combinations...  
Each trait, method and unit has to be identified if we want to share and reuse data



T1: Plant Height

M5: Highest pixel corresponding to plant

U3: pixel

Slide from L. Cabrera-Bosquet

# Ontologies, variables, descriptors

- Phenotyping variables
  - Crop Ontology
    - Collection of species specific ontologies
    - <https://www.croponontology.org/>
    - Plus <https://urgi.versailles.inra.fr/ontologyportal>
    - Methods can be specific of a community (CGIAR, INRAE)
  - Consortium dedicated ontology
  - Or
  - contribution to the Crop Ontologies
  - → Inclusion of the Crop ONTology triplat in MIAPPE
- Ontology term search engine
  - <https://www.ebi.ac.uk/ols4>

methods and scales English

MORPHO\_OIV\_225: Berry: color of skin **VARIABLE**

Ontology name	Vitis-intra-ontology
Identifier	CO_356:1000017
Name	MORPHO_OIV_225
Synonyms	Berry: color of skin
Institution	INRAE
Scientist	Eric Duchene
Date	00/00/00
Crop	VITIS

**Berry: color of skin** **TRAIT**

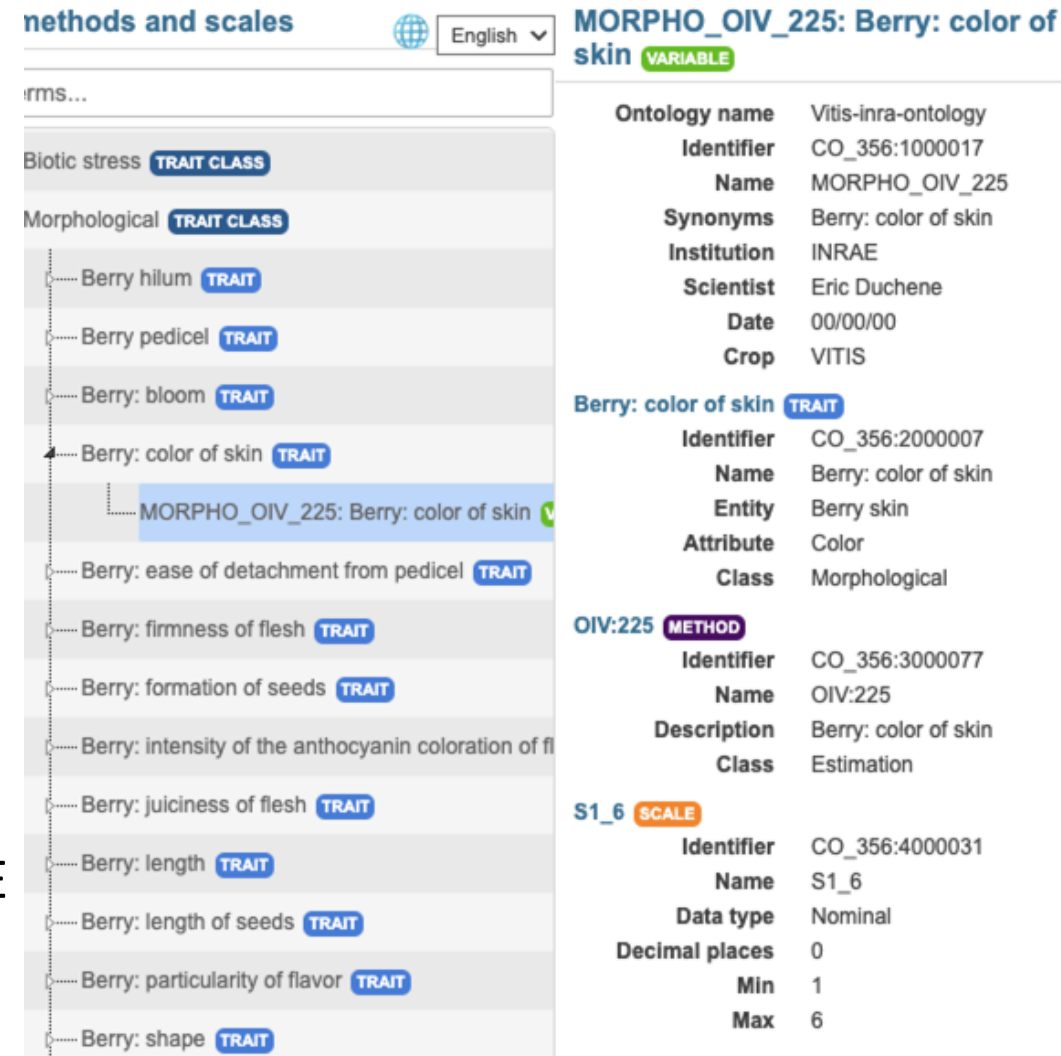
Identifier	CO_356:2000007
Name	Berry: color of skin
Entity	Berry skin
Attribute	Color
Class	Morphological

**OIV:225** **METHOD**

Identifier	CO_356:3000077
Name	OIV:225
Description	Berry: color of skin
Class	Estimation

**S1\_6** **SCALE**

Identifier	CO_356:4000031
Name	S1_6
Data type	Nominal
Decimal places	0
Min	1
Max	6



- **Phenotype Structure Standard**



## Minimal Information About Plant Phenotyping Experiments

[www.miappe.org](http://www.miappe.org)

- Many stakeholders
  - ◆ Elixir, Emphasis, Bioversity, North American PPN
- Open Community
  - ◆ Request for comments
  - ◆ Github Feature requests
  - ◆ Mailing lists
  - ◆ Meetings & Workgroups
  - ◆ 2nd version

- **Crops and woody plants**

Papoutsoglou *et al.* (2020) Enabling reusability and interoperability of plant phenomic datasets with MIAPPE 1.1. *New Phytol*, 227:260-273; <https://doi.org/10.1111/nph.16544>

MIAPPE					
line #	MIAPPE Check list	Definition	Example	Format	Cardinality
DM-1	<b>Investigation</b>	Investigations are research programmes with defined aims. They can exist at various scales (for example, they could encompass a grant-funded programme of work, the various components comprising a peer-reviewed publication, or a single experiment).			1 per MIAPPE submission
DM-2	<b>Investigation unique ID</b>	Identifier comprising the unique name of the institution/database hosting the submission of the investigation data, and the accession number of the investigation in that institution.	EBI12345678	Unique identifier	0-1
	<b>Investigation title</b>	Human-readable string summarising the investigation.	Adaptation of Maize to Temperate Climates: Mid-Density Genome-Wide Association Genetics and Diversity Patterns Reveal Key Genomic Regions, with	Free text (short)	1
Environment					
ENV-1	Non-exhaustive list of Environment Parameters.				
ENV-2	Environment parameters	Definition	Example	Format	
ENV-3	Growth facility				
ENV-4	<b>Air temperature</b>	List of hourly air temperature throughout the experiment.	22 °C	Numeric	
ENV-5	<b>Organ temperature</b>	List of hourly organ temperatures throughout the experiment.	18 °C	Numeric	
Experimental Factors					
TR-1	Non-exhaustive list of Experimental Factors that can be applied.				
TR-2	Factor type	Definition	Example factor values	Format	
TR-3	<b>Seasonal environment</b>	A plant treatment (EO:0001001) involving an exposure to a given conditions of regional seasons.	Spring season; dry season	Plant Environment Ontology:'EO_0007038'	
TR-4	<b>Air treatment regime</b>	The treatment involving an exposure to wind/air with varying degree of temperature, which may depend on the study type or the regional environment.	28/25°C ( Day/Night )	Plant Environment Ontology:'EO_0007161'	
TR-5	<b>Soil temperature regime</b>	A physical plant treatment (EO:0007316) involving an exposure to varying degree of temperature, which may depend on regional environment.	27/25°C ( Day/Night )	Plant Environment Ontology:'EO_0007161'	



- **Phenotype Structure Standard**



Minimum Information for Biological and Biomedical Investigations

A collection of the historical MIBBI foundry reporting guidelines. The minimum information standard is a set of guidelines for reporting data derived by relevant methods in biosciences. If followed, it ensures that the data can be easily verified, analysed and clearly

- **Biologist Friendly**
  - Clear definitions and examples
  - Excel templates
  - Trainings
- **Minimal and sufficient list of metadata:**
  - The objective of the experiment
  - Who contributed to the experiment
  - What were the experimental procedures
  - What was the biological material experimented
  - ...



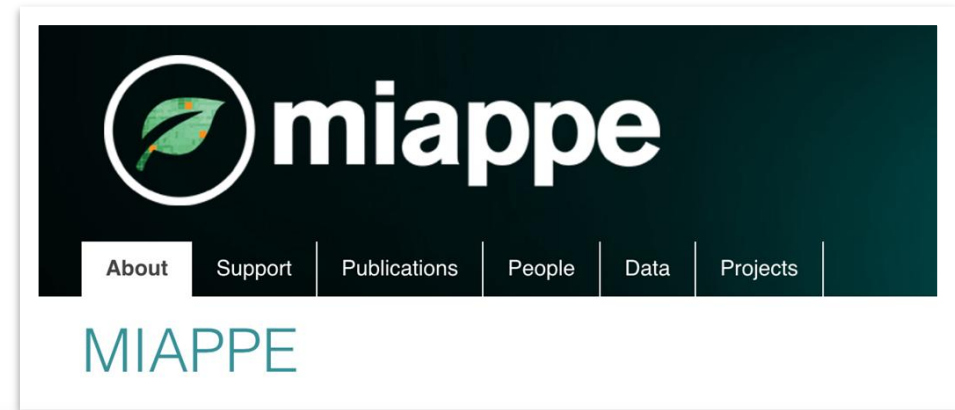
# • MIAPPE Specifications

- [www.miappe.org](http://www.miappe.org)

## MIAPPE primers

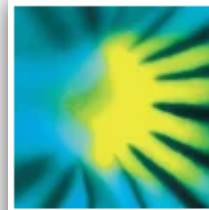
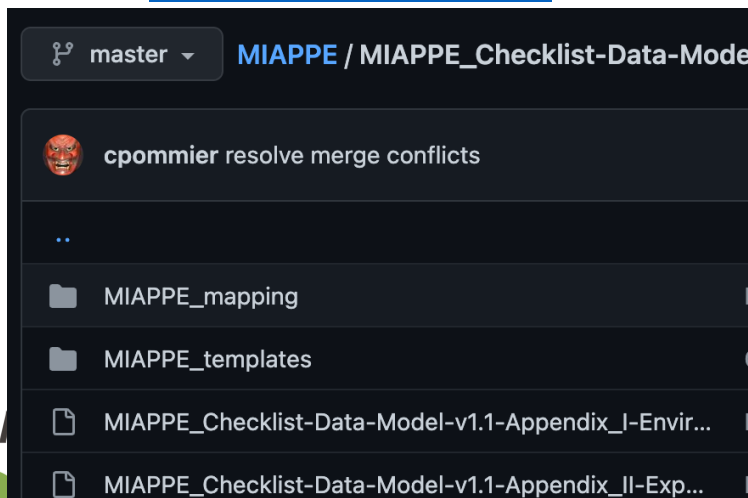
See the [support page](#) for full informations

- The latest specifications, [data model overview](#)
- The latest specifications [field list with description](#)



- Github

- [https://github.com/MIAPPE/MIAPPE/tree/master/MIAPPE\\_Checklist-Data-Model-v1.1](https://github.com/MIAPPE/MIAPPE/tree/master/MIAPPE_Checklist-Data-Model-v1.1)



## New Phytologist

Methods | [Open Access](#) |

### Enabling reusability of plant phenomic datasets with MIAPPE 1.1

Evangelia A. Papoutsoglou , Daniel Faria, Daniel Arend, Elizabeth Arnaud, Ioannis N. Athanasiadis, Inês Chaves, Frederik Coppens, Guillaume Cornut, Bruno V. Costa, Hanna Ćwiek-Kupczyńska, Bert Droesbeke, Richard Finkers, Kristina Gruden, Astrid Junker, Graham J. King, Paweł Krajewski, Matthias Lange, Marie-Angélique Laporte, Célia Michotey, Markus Oppermann, Richard Ostler, Hendrik Poorter, Ricardo Ramírez-Gonzalez, Živa Ramšak, Jochen C. Reif, Philippe Rocca-Serra, Susanna-Assunta Sansone, Uwe Scholz, François Tardieu, Cristobal Uauy, Björn Usadel, Richard G. F. Visser, Stephan Weise, Paul J. Kersey, Célia M. Miguel, Anne-Françoise Adam-Blondon, Cyril Pommier ... [See fewer authors](#) ^

First published: 14 March 2020 | <https://doi.org/10.1111/nph.16544> | Citations: 10

# • MIAPPE Specifications

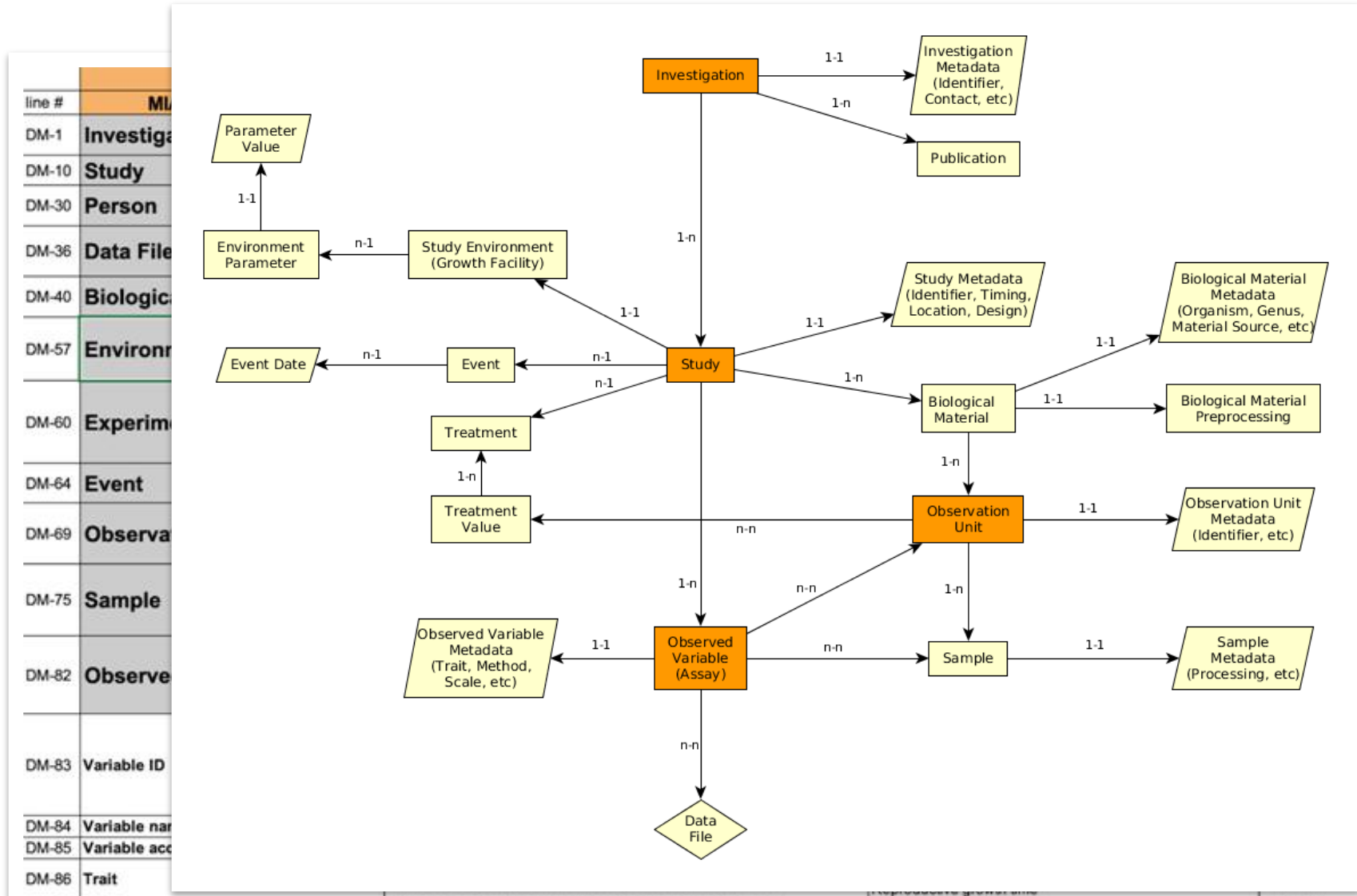
- Specification table
- Sections
- Metadata Fields

line #	MIAPPE Check list	MIAPPE			Cardinality
		Definition	Example	Format	
DM-1	<b>Investigation</b>	Investigations are research programmes with defined aims. They can exist at various scales (for example, they could encompass a grant-funded programme of work, the various components comprising a peer-reviewed publication, or a single experiment).			1 per MIAPPE submission
DM-10	<b>Study</b>	A study (or experiment) comprises a series of assays (or measurements) of one or more types, undertaken to answer a particular biological question.			1+ per investigation
DM-30	<b>Person</b>	A human involved in the investigation or specifically any of its studies.			1+ per investigation / 0+
DM-36	<b>Data File</b>	A file or digital object holding observation data recorded during one or more assays of the study, typically in tabular form. Multiple data files may be provided per study, and each file can include observations for several observation units and several observed variables.			0+ per study
DM-40	<b>Biological Material</b>	The biological material being studied (e.g. plants grown from a certain bag or seed, or plants grown in a particular field). The original source of that material (e.g., the seeds or the original plant cloned) is called the material source, which, when held by a material repository, should have its stock identified.			1+ per study; 0+ per observation unit
DM-57	<b>Environment</b>	Environmental parameters that were kept constant throughout the study and did not change between observation units or assays. Environment characteristics that vary over time, i.e. environmental variables, should be recorded as Observed Variables (see below).			0-1 per study
DM-60	<b>Experimental Factor</b>	The object of a study is to ascertain the impact of one or more factors on the biological material. Thus, a factor is, by definition a condition that varies between observation units, which may be biotic (pest, disease interaction) or abiotic (treatment and cultural practice) in nature. Depending on the level of the data, an experimental factor can be either "what is the factor applied to the plant" (i.e. Unwatered), or the "environmental characterisation" (i.e. if no rain on unwatered plant : Drought ; if rain on unwatered plant: Irrigated)			0+ per study; 0+ per observation unit
DM-64	<b>Event</b>	An event is discrete occurrence at a particular time in the experiment (which can be natural, such as rain, or unnatural, such as planting, watering, etc). Events may be the realization of Factors or parts of Factors, or may be confounding to Factors. Can be applied at the whole study level or to only a subset of observation units.			0+ per study/observation unit
DM-69	<b>Observation Unit</b>	Observation units are objects that are subject to instances of observation and measurement. An observation unit comprises one or more plants, and/or their environment. There can be pure environment observation units with no plants. Synonym: Experimental unit.			1+ per study
DM-75	<b>Sample</b>	A sample is a portion of plant tissue harvested, non-harvested or extracted from an observation unit for the purpose of sub-plant observations and/or molecular studies. A sample must be used when there is a physical sample that needs to be stored and traced. Otherwise, observations made at the sub-plant level should be recorded as plant level observations using the observed variables to characterize the object of the observation (e.g. Berry sugar content, Fruit weight, Grain Protein content, Leaf 1 width, Leaf 2 width, Leaf 2 length).			0+ per observation unit
DM-82	<b>Observed Variable</b>	An observed variable describes how a measurement has been made. It typically takes the form of a measured characteristic of the observation unit (plant or environmental trait), associated to the method and unit of measurement. Multiple variables with the same combination of trait, method and scale can be used in association with different plant parts (leaf 1, leaf 2), when this distinction is necessary for observations referring to different parts of the same observation unit.			1+ per study
DM-83	<b>Variable ID</b>	Code used to identify the variable in the data file. We recommend using a variable definition from the Crop Ontology where possible. Otherwise, the Crop Ontology naming convention is recommended: <trait abbreviation>_<method abbreviation>_<scale abbreviation>. A variable ID must be unique within a given investigation.	Ant_Cmp_Cday	Unique identifier	1
DM-84	<b>Variable name</b>	Name of the variable.	Anthesis computed in growing degree days	Free text	0-1
DM-85	<b>Variable accession number</b>	Accession number of the variable in the Crop Ontology	CO_322:0000794	Crop Ontology term	0-1
DM-86	<b>Trait</b>	Name of the (plant or environmental) trait under observation	Anthesis time Reproductive growth time	Free text	1
DM-87	<b>Trait accession number</b>	Accession number of the trait in a suitable controlled vocabulary (Crop Ontology, Trait Ontology).	CO_322:0000030 TO:0000366	Term from Plant Trait Ontology, Crop Ontology, or XML Environment Ontology	0-1



# MIAPPE Specifications

- Specification table
- Sections
- Metadata Fields
- Linked and organized



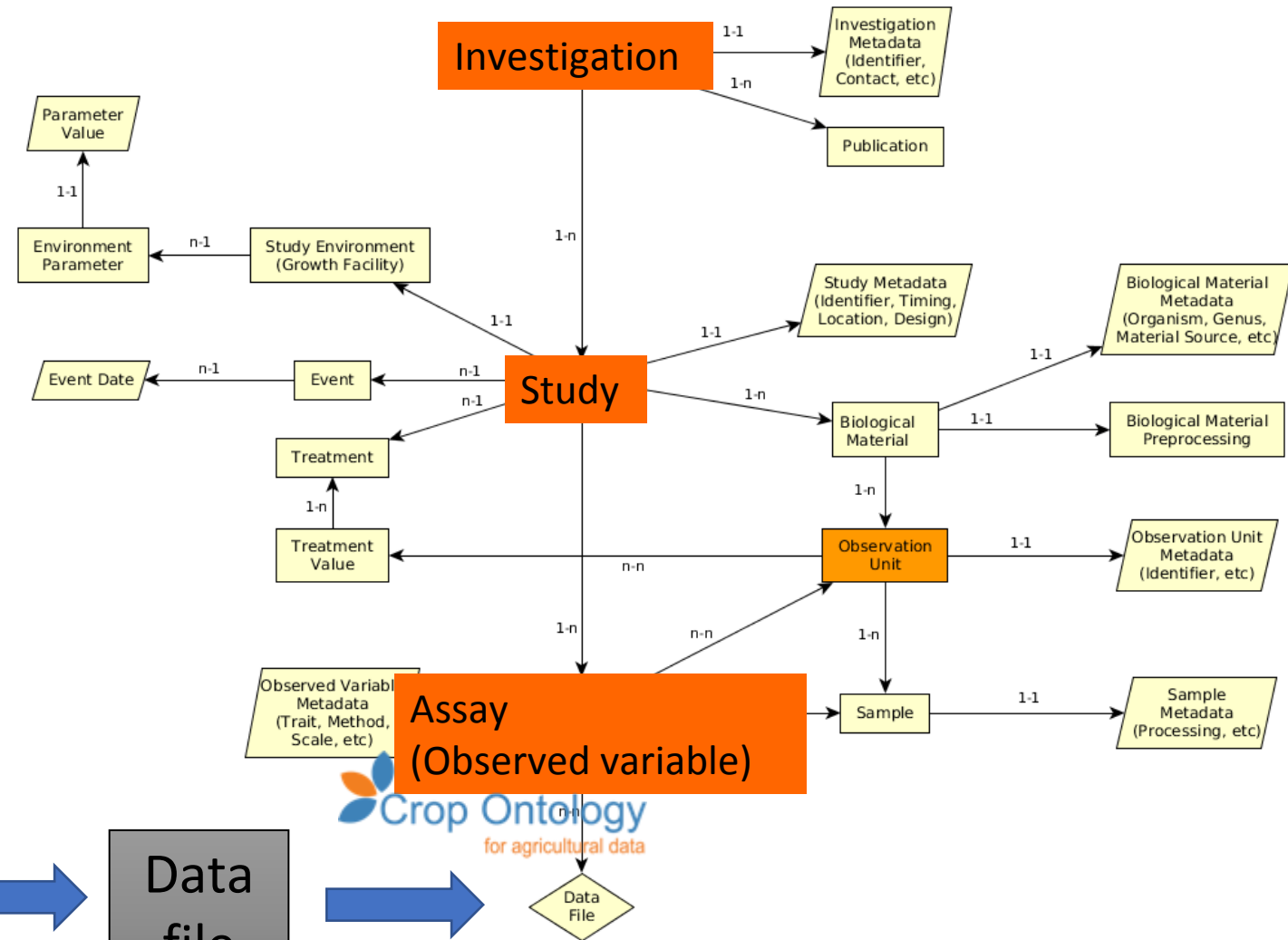
line #	Metadata Field
DM-1	Investigation
DM-10	Study
DM-30	Person
DM-36	Data File
DM-40	Biological Material
DM-57	Environment
DM-60	Experiment
DM-64	Event
DM-69	Observation Unit
DM-75	Sample
DM-82	Observed Variable
DM-83	Variable ID
DM-84	Variable name
DM-85	Variable accession number
DM-86	Trait

Format	Cardinality
name of work, the	1 per
	1+ per
	1+ per
vided per study, and	0+ per
aterial (e.g., the seeds or	1+ per
	0+ per
eristics that vary over	0-1 per
tween observation	0+ per
mental factor can be	0+ per
in on unwatered plant:	0+ per
Events may be the	0+ per
nits.	0+ per
their environment.	1+ per
	1+ per
olecular studies. A	0+ per
be recorded as plant	0+ per
nt, Leaf 1 width, Leaf 2	0+ per
int or environmental	1+ per
iation with different	1+ per
	1
	0-1
term	0-1
	1

DM-87	Trait accession number	Accession number of the trait in a suitable controlled vocabulary (Crop Ontology, Trait Ontology).	CO_322:0000030 TO:0000366	Term from Plant Trait Ontology, Crop Ontology, or XML Environment Ontology	0-1
-------	------------------------	--	------------------------------	--	-----

# MIAPPE V1.1 Overview the (ISA) backbone

- **Investigation:** whole dataset
  - field, greenhouse,
  - One to many location
  - One to many years
- **Study** : one experiment in one location for one to several year
- **Assay:**
  - trait or indice (Pheno or Env) observed
  - Level + Trait + Method + Scale/Unit
- **Level:**
  - Plant
  - Microplot
  - Block
  - Trial/Study





# Data management tools

- Data acquisition & management during the experiments
- Simple phenotyping network
  - Fairdom Seek (AGENT H2020)
  - Simple shared space for data, files SOP, etc..
  - Avoid “sent by email”
  - **Usable beyond Phenotyping**
- Fully instrumented Phenotype
  - Emphasis solutions (PIPPA, PHIS, IPK, ...)
    - <https://emphasis.plant-phenotyping.eu/services/emphasis-pilots/data-services>
  - **Upcoming BigData platform Cor practice presentation on PHIS**

**WP3-Phenotypic-Historical**

Overview | Related items

Please use the following template: [Eurisco Phenotyping Historical data template](#)

See the [Helpdesk FAQ](#) for information on data management in FAIRDOM.

The three required traits are:

- plant height, written "Plant height"
- flowering time, either as "Days to heading" (Type: Measurement), or "Date of heading" (Type: Date)
- thousand-kernel weight, written "Thousand Kernel Weight"

SEEK ID: <https://urgi.versailles.inrae.fr/fairdom/investigations/2>

Projects: [H2020-AGENT](#)

Investigation position: 4

**Selected:** [WP3-Phenotypic-Historical](#) (Investigation)

**Description:** Please use the following template: [Eurisco Phenotyping Historical data...]

**SEEK ID:** <https://urgi.versailles.inrae.fr/fairdom/investigations/2>

- WP3-Phenotypic-Historical
  - CREA-CI-Phenotypic Historical Study wheat
    - CREA\_Phenotypic historical data wheat\_01-2023
      - CREA Phenotypic hystorical assay\_01-2023
      - CREA Phenotypic hystorical assay\_01-2023 (Validation Report v4)
  - CRI-Phenotypic-Historical-Wheat
    - CRI-Phenotypic-Historical-Wheat

**PHENOME**

- Scientific Organization
- Organizations
- Projects
- Devices
- Shared resources
- Variables

**Germplasm**  
Manage Genetic Resources Information

Germplasm | Germplasm Group

Type: Select a germplasm type

Species: Select a species

Production year

Selected Germplasm 0 | Display | Action

Showing 0 to 20 of 6,277 entries

<input type="checkbox"/>	Name	Type
<input type="checkbox"/>	10-108_ACN_15-2	Variety

# Data repositories

- Sustainable data access
- Generic
  - Zenodo
  - Dataverse
  - With MIAPPE additions
  - Easier to use
  - Training : <https://elixir.mf.uni-lj.si/increasing23> (TESS soon)
- Dedicated to Phenotyping
  - PHIS
  - PIPPA
  - BMS
  - BreedBase
  - ....
  - More complex
  - Complementary to generic repositories
  - Generic datasets can be linked to specific database
    - Best of the two worlds
    - <https://doi.org/10.15454/IASSTN>

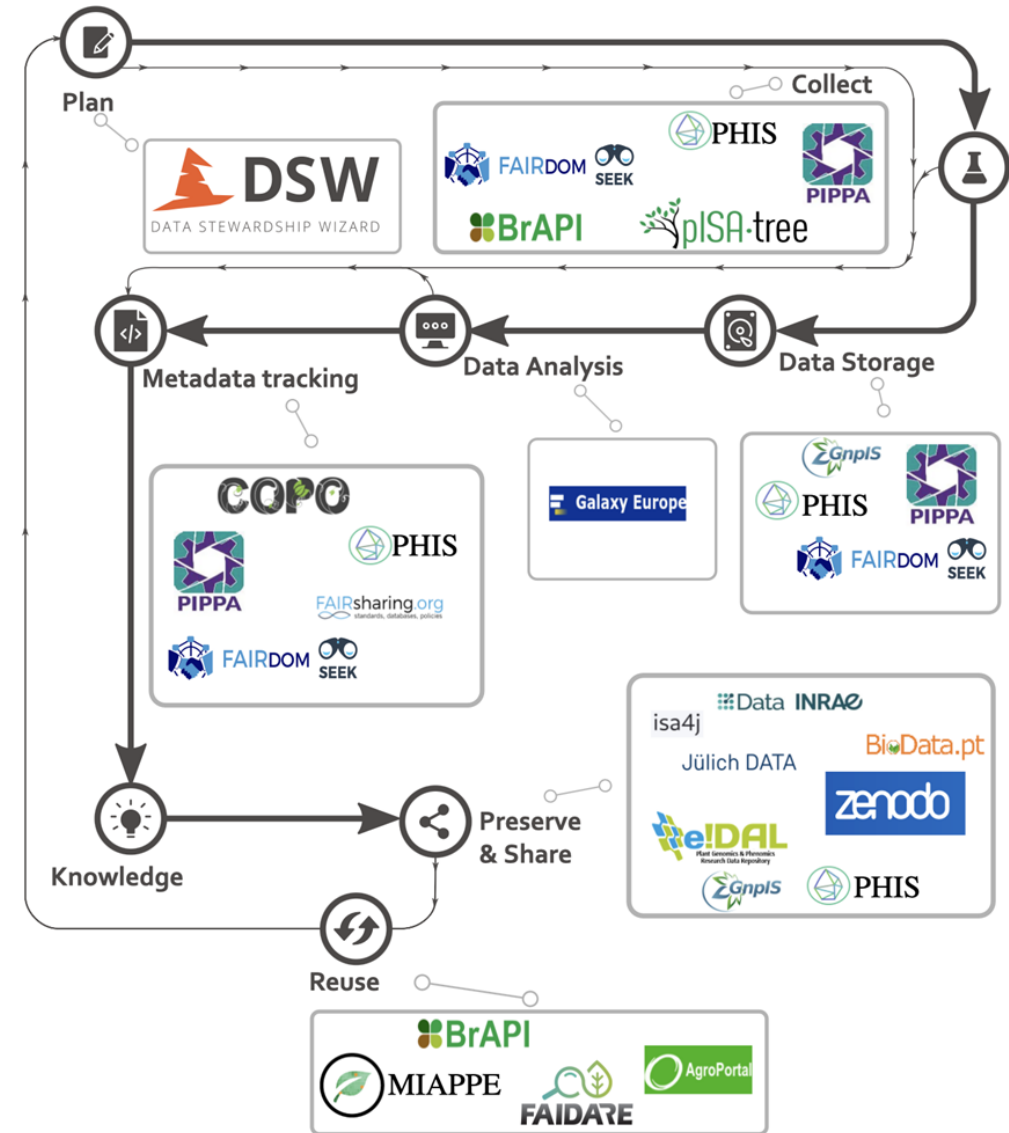


Figure 1. The plant phenomics tool assembly.

# • Data Integration between silos, From Phenotyping to Genotyping

## Phenomics



Biological Material

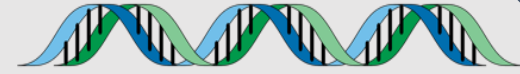
Field Experiment

Location



Identifying key resources/pivot objects

## Genetics Genomics Omics



Biological Material

Samples

Omic Experiment



## Environment



Location



# • Data Integration between silos, From Phenotyping to Genotyping

Phenomics



Biological Material

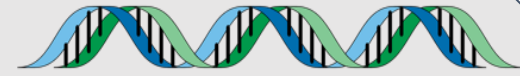
Field Experiment

Location

Interoperability pivot  
Key shared resources



Genetics  
Genomics  
Omics



Biological Material

Samples

Omic Experiment

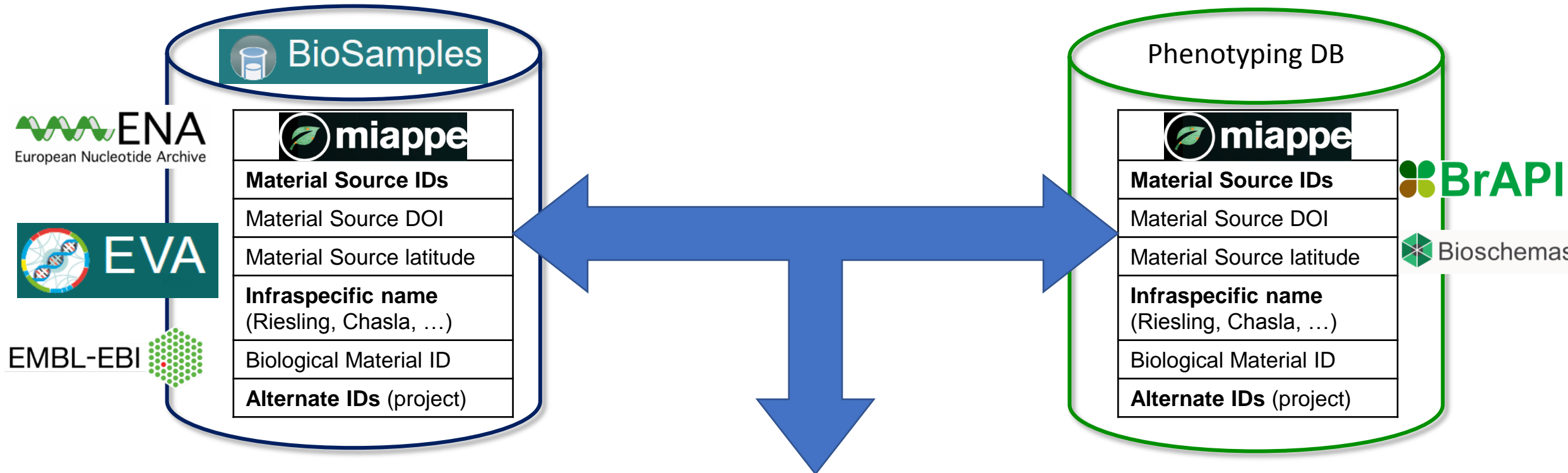


Environment



Location

# Data Integration between silos, From Phenotyping to Genotyping



URGI Data providers More... <https://urgi.versailles.inrae.fr/faidare/> elixir

### FAIR Data-finder for Agronomic REsearch

Search keywords

**Germplasm** Trait Reset all

**Crops** (common name, species, genus, subtaxa & synonyms) Search crops

**Germplasm list** (panel, collection & population) Search germplasm lists

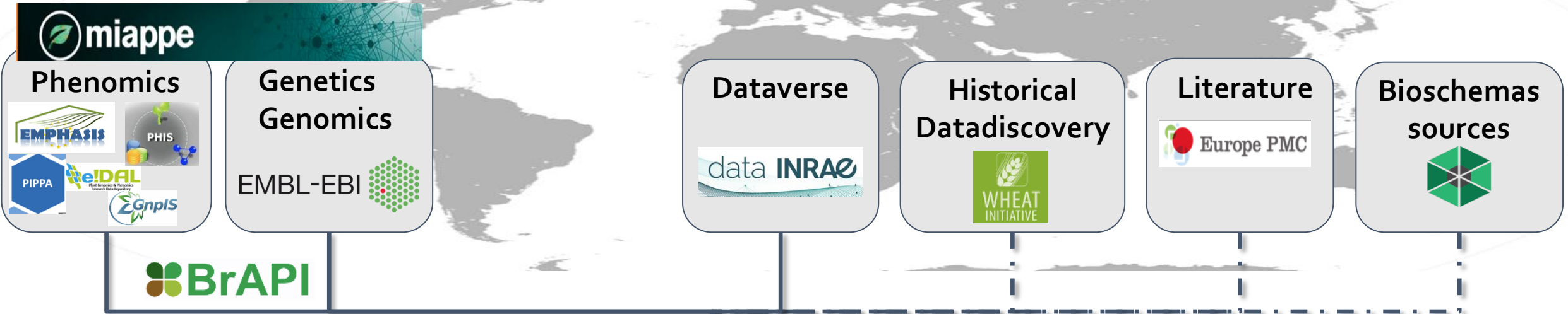
**Sources**

- URGI GnpIS (81,335)
- EBI European Nucleotide Archive (44,975)
- CIRAD TropGENE (722)
- VIB PIPPA (692)
- IBET BioData (67)
- IWGSC@GnpIS (18,814,632)
- Evoltree@GnpIS (5,354)
- OpenMinTeD@GnpIS (3,392)
- EBI Ensembl Plants

Community data discovery portals



# Findability Global Data discovery portal



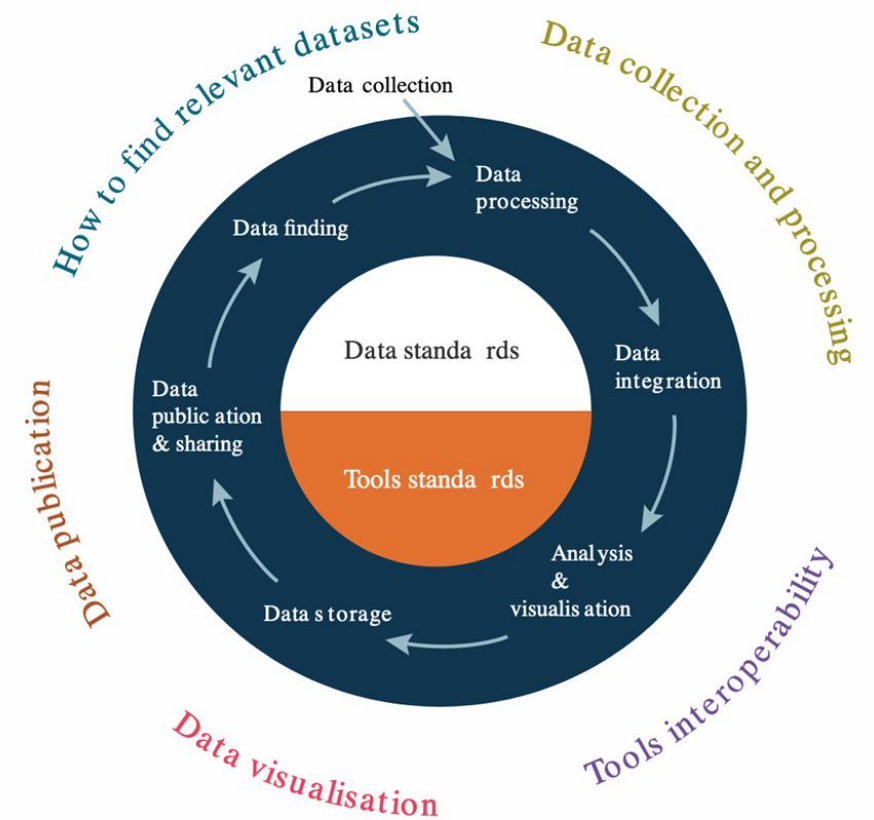
<https://urgi.versailles.inrae.fr/aidare/>

Full text  
+  
Fine criteria  
+  
Link back



# Take Home Message

- Data management relies on a complex lifecycle
  - Observation & Measures
  - Deriving and reducing data
  - Linking different datasets
- All steps must be defined
  - data management plan
- Not all step of data must be shared
- Raw and final data should be shared
- With sufficient provenance
- Community driven
  - [https://rdmkit.elixir-europe.org/plant\\_sciences](https://rdmkit.elixir-europe.org/plant_sciences)
- Data standard and tools are easier to use
- Data acquisition tools are often community specific, but reusable
  - PHIS
  - CGIAR tools
- Open science (policy): Sustainable sharing and Findability are keys (Data tombs effect)



# Aknowledgments



## H2020 AGENT

N. Stein (IPK, coord), P. Kersey (RBGK), M. Alaux (INRAE), S. Weise (IPK), C. Pommier (INRAE), M. Lange (IPK), R. Finkers (WUR), J. Destin (INRAE)

## Elixir Plant community & platforms



Beier S., Gruden C., Pommier C., Coppens F, Scholz U., Lange M., Contreras B., Adam Blondon AF, Faria D, Chavez I, Miguel C, Droedsbek B, Finkers R, Papoutsoglou E, Olster R, Ramsak Z, ...

## MIAPPE community



ELIXIR Plant Community, Krajewsky P, Cwiek H, Tardieu F, Usadel B, Arend D, Arnaud E, Junker A, King G, Laporte MA, Poorter H, Reif J, Rocca-Serra P, Sansone SA, Kersey P, And many more!



### Breeding API

Selby P, Mueller L, Robbins K, Backlund JE, ... , And many more!

## Crop Ontology



Arnaud E, Laporte MA, ...

## Emphasis



Tardieu F, Usadel B, Arend D, Junker A, Poorter H, Neveu P, Pierushka R, Shur U... And many more!



EMPHASIS



EMBL-EBI



ant Phenomics data management challenges and s ed on

October 2023 / Cyril Pommier